

Article

LSSVR Model of G-L Mixed Noise-Characteristic with Its Applications

Shiguang Zhang ^{1,2,3,*} , Ting Zhou ^{4,*}, Lin Sun ^{1,3} , Wei Wang ¹ and Baofang Chang ¹

¹ College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China; sunlin@htu.edu.cn (L.S.); wangwei@htu.edu.cn (W.W.); changbaofang@htu.edu.cn (B.C.)

² School of Computer Science and Technology, Tianjin University, Tianjin 300350, China

³ Engineering Lab of Intelligence Business and Internet of Things, Xinxiang 453007, China

⁴ The State-Owned Assets Management Office, Henan Normal University, Xinxiang 453007, China

* Correspondence: zhangshiguang@htu.edu.cn (S.Z.); zhouting@htu.edu.cn (T.Z.); Tel.: +86-373-332-6271 (S.Z.)

Received: 27 April 2020; Accepted: 2 June 2020; Published: 6 June 2020



Abstract: Due to the complexity of wind speed, it has been reported that mixed-noise models, constituted by multiple noise distributions, perform better than single-noise models. However, most existing regression models suppose that the noise distribution is single. Therefore, we study the Least square SVR of the Gaussian–Laplacian mixed homoscedastic ($GLM - LSSVR$) and heteroscedastic noise ($GLMH - LSSVR$) for complicated or unknown noise distributions. The ALM technique is used to solve model $GLM - LSSVR$. $GLM - LSSVR$ is used to predict short-term wind speed with historical data. The prediction results indicate that the presented model is superior to the single-noise model, and has fine performance.

Keywords: Least square SVR; Gaussian–Laplacian mixed noise-characteristic; empirical risk loss; equality constraint; wind-speed forecasting

1. Introduction

In practical applications, if the data are collected in a multi-source environment, the noise distribution is complex and unknown. Therefore, it is almost impossible for a single-noise distribution to clearly describe the real-noise [1]. $LSSVR$ is a method of LR that implements a sum-of-squares error function together with regularization, thus controlling the bias–variance trade-off [2,3]. It is intended to find the concealed linear structures in the original data [4,5]. For the sake of transition from linear to nonlinear function, the following generalization can be made [6]: by mapping input vectors into a high-dimensional feature space H (H is Hilbert space) through some nonlinear-mapping, seek the solution of the optimization problem in space H . Using a suitable kernel function $K(\bullet, \bullet)$, nonlinear-mappings can be estimated by kernel $LSSVR$, which is an extended LR with kernel techniques. In recent years, $LSSVR$ as a data-rich nonlinear forecasting tool has been increasingly welcomed [7], which is applicable in many different contexts [8–10], such as machine learning, optical character recognition, and especially wind speed/power forecasting.

Generally, the existing techniques used for wind-speed forecasting include: (i) physical; (ii) statistical (also called data-driven); and (iii) artificial intelligence (AI)-based methods. The physical models attempt to estimate wind flow around and inside the wind farm using physical laws governing the atmospheric behavior [11,12]. The statistical models seek the relationships between a set of explanatory variables and the on-line measured generation data, and the historical wind speed data recorded at the site are only used to establish the statistical model. We can model it in a variety of ways, including persistence method

and auto-regressive model [13,14]. AI methods include artificial neural networks (ANNs) [15], deep learning [16], SVR machines [17,18], and the hybrid methods [19,20].

Suykens et al. [21–23] proposed least square support vector regression model with Gaussian noise (*LSSVR*, also known as kernel ridge regression (*KRR*)). Mixed-model based on multi-objective optimization [24,25], mixed-method based on singular spectrum analysis, firefly algorithm, and BP neural network predict wind speed with complicated noise [26], indicating that the mixed prediction method has the ability of powerful prediction. Mixed *LSSVR* machine [27] is applied to forecast the wind speed noise, which improves performance of wind-speed prediction. *GLM – SVR* [28] models fitted by Gaussian–Laplacian (G-L) mixed noise are developed, and good performance is obtained compared with the existing regression algorithm.

To solve the above problems, we study model *LSSVR* of G-L mixed noise-characteristic for complex or unknown noise distribution. In this case, we construct a technique to search the optimal solution of the corresponding regression task. Although many *LSSVR* algorithms have been implemented in past years, we exploit ALM method, as shown in Section 4. If the task is not differentiable or discontinuous, the sub gradient descent method can be employed, or the SMO [29] can also be used if there is a very large sample size.

The structure of this paper is as follows. Section 2 derives the optimal empirical risk loss by Bayesian principle. Section 3 constructs the *LSSVR* model of G-L mixed noise. Section 4 gives the solution and algorithm design of *GLM – LSSVR*. In Section 5, the numerical experiment of short-term wind-speed prediction is presented. Finally, we conclude the work.

2. Bayesian Principle to Mixed Noise Empirical Risk Loss

Given the Dataset

$$D_N = \{(A_1, y_1), (A_2, y_2), \dots, (A_N, y_N)\}, \tag{1}$$

where $A_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \in R^n, y_i \in R (i = 1, 2, \dots, N)$ is the training data. R represents real number set, R^n is the n -dimensional Euclidean-Space, and N is the sample size. Superscript T is the transpose of matrix. Assuming that the sample of dataset D_N is generated by the additive noise function ζ , the relationship between the measured value y_i and predicted value $f(A_i)$ is:

$$y_i = f(A_i) + \zeta_i, i = 1, 2, \dots, N \tag{2}$$

where ζ_i is random, i.i.d. (independent, identical probability distribution) with $p(\zeta_i)$ of mean μ and standard deviation σ . Generally, the noise *PDF* (probability density function) $p(\zeta) = p(y - f(A))$ is unknown. It is necessary to predict unknown target $f(A)$ from training set $D_f \subseteq D_N$.

Following the authors of [30,31], the optimal empirical risk loss in the sense of Maximum Likelihood (MLE) is

$$l(\zeta) = l(A, y, f(A)) = -\text{Log}p(y - f(A)), \tag{3}$$

i.e., the empirical risk loss $l(\zeta)$ is the log-likelihood of noise characteristic.

It is assumed that noise in Equation (2) is Laplacian, with *PDF* $p(\zeta) = \frac{1}{2}e^{-|\zeta|}$. By Equation (3), in MLE the optimal empirical risk loss should be $l(\zeta) = |\zeta|$.

Suppose noise in Equation (2) is Gaussian of zero mean and homoscedastic standard deviation σ . By Equation (3), the empirical risk loss of Gaussian noise with homoscedasticity is $l(\zeta) = \frac{1}{2\sigma^2}\zeta^2$. The noise in Equation (2) is Gaussian of zero mean and heteroscedastic standard deviation σ_i . By Equation (3), the empirical risk loss for Gaussian-noise with heteroscedasticity is $l(\zeta_i) = \frac{1}{2\sigma_i^2}\zeta_i^2 (i = 1, \dots, N)$.

Assume noise ζ in Equation (2) is the mixed noise of two kinds of noise with the *PDFs* $p_1(\zeta)$ and $p_2(\zeta)$, respectively. Suppose that $p(\zeta) = [p_1(\zeta)^{\lambda_1}] \cdot [p_2(\zeta)^{\lambda_2}]$. By Equation (3), the corresponding empirical risk loss of mixed-noise is

$$I(\xi) = \lambda_1 \cdot l_1(\xi) + \lambda_2 \cdot l_2(\xi). \tag{4}$$

where $l_1(\xi) > 0, l_2(\xi) > 0$ are the convex empirical risk losses of the above two kinds of noise characteristic, respectively. The weight factors are $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Figure 1 displays the Gaussian–Laplacian (G-L) empirical risk loss of different parameters (the parameter lambda is λ) [29].

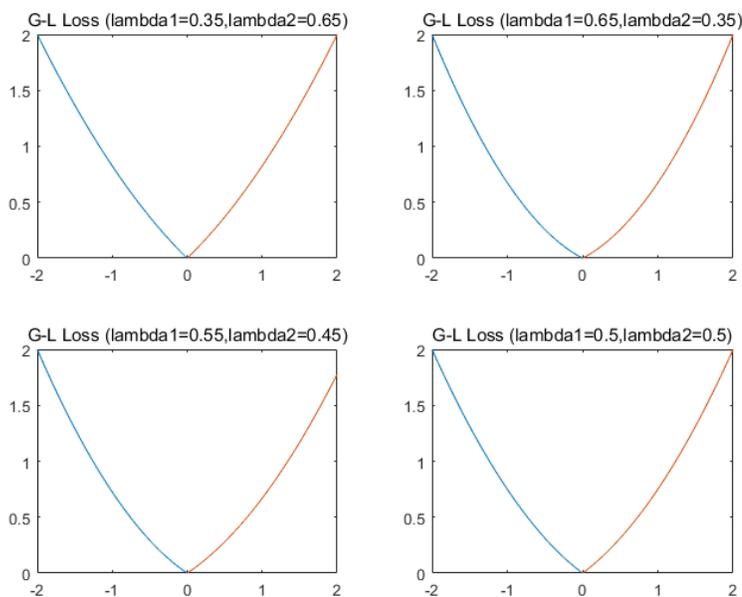


Figure 1. G-L empirical risk loss of different parameters.

3. LSSVR Model of G-L Mixed Noise-Characteristic

Given the training samples $D_f \subseteq D_N$, construct the linear regressor $f(A) = \omega^T \cdot A + b$. To deal with nonlinear problems, it can be summarized as follows: mapping input vectors $A_i \in R^n$ into high-dimension feature space H through the nonlinear mapping Φ (take a prior distribution), induced by nonlinear kernel function $K(A_i, A_j)$, kernel mapping Φ is any positive definite Mercer kernel.

Definition 1 ([6,28]). *Positive definite Mercer kernel: Assume that X is a subset of R^n . Assume that the kernel function $K(A_i, A_j)$ defined on $X \times X$ is a positive definite Mercer kernel function the kernel mapping Φ is called a positive definite Mercer kernel if there is mapping $\Phi : X \rightarrow H$ (H is Hilbert Space), such that*

$$K(A_i, A_j) = (\Phi(A_i) \cdot \Phi(A_j)), (i, j = 1, 2, \dots, N). \tag{5}$$

where (\cdot) represents the inner-product in Space H .

Therefore, the optimization problem of Space H is solved. At present, the input vectors $(A_i \cdot A_j)$ are replaced by inner product $(\Phi(A_i) \cdot \Phi(A_j))$ in feature space H . Through the use of kernel $K(A_i, A_j) = (\Phi(A_i) \cdot \Phi(A_j))$, the linear model be extended to a nonlinear LSSVR.

In general, the mixed distribution has fine approximation ability to any continuous distribution. When there is no prior knowledge of real-noise, it can well adapt to unknown or complicated noise. Thus,

it is presented that a uniform model *LSSVR* with mixed noise characteristics (*M – LSSVR*). The primal problem of model *M – LSSVR* is formalized as

$$\begin{aligned} \text{Min}\{g_{P_{M-LSSVR}} &= \frac{1}{2}\omega^T \cdot \omega + \frac{C}{N} \cdot [\lambda_1 \cdot \sum_{i=1}^N (l_1(\xi_i)) + \\ &\lambda_2 \cdot \sum_{i=1}^N (l_2(\xi_i))]\} \\ \text{s.t. : } \xi_i &= y_i - \omega^T \cdot \Phi(A_i) - b \end{aligned} \tag{6}$$

where parameter $\omega \in R^n$ represents weight-vector, b is the bias-term, $C > 0$ is the penalty parameter, and the weight factors are $\lambda_1, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1$. $(A_i, y_i) \in D_N, \Phi(A)$ is a nonlinear mapping which transfers the input dataset to a higher-dimensional feature space H . $\xi_i = y_i - \omega^T \cdot \Phi(A_i) - b$ is the random noise variable at time $i (i = 1, 2, \dots, N)$. $l_1(\xi_i) > 0, l_2(\xi_i) > 0 (i = 1, 2, \dots, l)$ is the convex loss-functions for noise characteristic in sample-point $(A_i, y_i) \in D_N ((i, j = 1, 2, \dots, N))$.

In the application domain, most distributions do not obey Gaussian distribution, and they also do not satisfy Laplacian distribution. the noise distribution is complicated, and it is almost impossible to describe real noise with a single distribution. It has been reported that mixed noise models, constituted by multiple noise distributions, perform better than single-noise model [1]. As the function fitting -machine, the goal is to estimate an unknown function $f(A)$ from dataset $D_f \subseteq D_N$. In this section, G-L mixed homoscedastic and heteroscedastic noise distributions are used to fit complicated noise characteristic.

3.1. LSSVR Model of G-L Mixed Homoscedastic Noise-Characteristic

Suppose noise in Equation (2) is Gaussian of zero mean and homoscedastic standard deviation σ . By Equation (3), we have that the empirical risk loss of homoscedastic-Gaussian-noise characteristic is $l_1(\xi) = \frac{1}{2\sigma^2} \cdot \xi^2$. The Laplacian-noise is $l_2(\xi) = |\xi|$. Adopting G-L mixed homoscedastic noise distribution to fit complicated noise-characteristic, by Equation (4), the empirical risk loss about G-L mixed homoscedastic noise is $l(\xi) = \frac{\lambda_1}{2\sigma^2} \cdot \xi^2 + \lambda_2 \cdot |\xi|$. Putting forward the *LSSVR* model of G-L mixed homoscedastic noise-characteristic (*GLM – LSSVR*), the primal problem of *GLM – LSSVR* is depicted as

$$\begin{aligned} \text{Min}\{g_{P_{GLM-LSSVR}} &= \frac{1}{2}\omega^T \cdot \omega + \frac{C}{N} \cdot (\frac{\lambda_1}{2\sigma^2} \cdot \sum_{i=1}^N \xi_i^2 \\ &+ \lambda_2 \cdot \sum_{i=1}^N |\xi_i|)\} \\ \text{s.t. : } \xi_i &= y_i - \omega^T \cdot \Phi(A_i) - b \end{aligned} \tag{7}$$

where parameter vector $\omega \in R^n, \sigma^2$ is homoscedastic, $C > 0$ is a penalty parameter, and the weight factors are $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Proposition 1. *The solution of the primal problem in Equation (7) of GLM – LSSVR is existent and unique about ω .*

Theorem 1. *The dual problem of the primal problem in Equation (7) is*

$$\begin{aligned} \text{Max}\{g_{D_{GLM-LSSVR}} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \cdot \alpha_j \cdot K(A_i, A_j) + \\ &\sum_{i=1}^N \alpha_i \cdot y_i - \frac{N}{2C \cdot \lambda_1} \sum_{i=1}^N (\sigma^2 \cdot \alpha_i - C \cdot \lambda_2)^2\} \\ \text{s.t. : } \sum_{i=1}^N \alpha_i &= 0 \end{aligned} \tag{8}$$

where σ^2 is homoscedastic, $C > 0$ is a penalty parameter, and the weight factors are $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Proof. We introduce Lagrange functional $L(\omega, b, \alpha, \xi)$ as $L(\omega, b, \alpha, \xi) = \frac{1}{2}\omega^T \cdot \omega + \frac{C}{N} \cdot (\frac{\lambda_1}{2\sigma^2} \cdot \sum_{i=1}^N \xi_i^2 + \lambda_2 \cdot \sum_{i=1}^N \xi_i) + \sum_{i=1}^N \alpha_i(y_i - \omega^T \cdot \Phi(A_i) - b - \xi_i)$.

Minimizing $L(\omega, b, \alpha, \xi)$ and deriving the partial-derivative ω, b, ξ , respectively, on the basis of KKT-conditions, we get

$$\nabla_{\omega}(L) = 0, \nabla_b(L) = 0, \nabla_{\xi}(L) = 0.$$

We obtain

$$\omega = \sum_{i=1}^N \alpha_i \cdot \Phi(A_i),$$

$$\sum_{i=1}^N \alpha_i = 0,$$

$$\frac{C}{N} \cdot (\frac{\lambda_1}{\sigma^2} \cdot \xi_i + \lambda_2) - \alpha_i = 0 (i = 1, 2, \dots, N).$$

The extreme condition is replaced by $L(\omega, b, \alpha, \xi)$, and the maximum value of α is obtained. The dual problem in Equation (8) of the primal problem in Equation (7) is derived. □

Therefore,

$$\omega_i = \sum_{i=1}^N \alpha_i \cdot \Phi(A_i),$$

$$b = \frac{1}{N} \sum_{i=1}^N [y_i - \sum_{j=1}^N \alpha_j \cdot K(A_i, A_j) - \frac{1}{\lambda_1} \cdot (\frac{N \cdot \sigma^2 \cdot \alpha_i}{C} - \lambda_2)].$$

The decision-maker for *GLM – LSSVR* may be represented as

$$f(A) = \omega^T \cdot \Phi(A) + b = \sum_{i=1}^N \alpha_i K(A_i, A) + b.$$

where the parameter vector $\omega \in R^n$, $\Phi : R^n \rightarrow H$, $(\Phi(A_i) \cdot \Phi(A_j))$ is the inner-product of H and $K(A_i, A_j) = (\Phi(A_i) \cdot \Phi(A_j))$ is the kernel-function.

Suppose the noise in Equation (2) is Gaussian homoscedastic noise, which is Gaussian noise of zero mean and the homoscedastic variance σ^2 . Thus, the dual problem of *LSSVR* can be derived by Theorem 2:

$$\begin{aligned} \text{Max}\{g_{D_{LSSVR}} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i \cdot \alpha_j \cdot K(A_i, A_j)) \\ &+ \sum_{i=1}^N (\alpha_i \cdot y_i) - \frac{N}{2C} \cdot \sum_{i=1}^N (\sigma^2 \cdot \alpha_i^2)\} \\ D_{GN-KRR} : \text{s.t.} \quad &\sum_{i=1}^N \alpha_i = 0. \end{aligned} \tag{9}$$

3.2. LSSVR Model of G-L Mixed Heteroscedastic Noise-Characteristic

It is assumed that the noise in Equation (2) is Gaussian of zero mean and heteroscedastic standard deviation σ_i , that is $\sigma_i \neq \sigma_j, i \neq j (i, j = 1, \dots, N)$. From Equation (3), the empirical risk loss of heteroscedastic Gaussian-noise characteristic is $l_1(\xi_i) = \frac{1}{2\sigma_i^2} \cdot \xi_i^2$ and the loss-function of Laplacian-noise is $l_2(\xi_i) = |\xi_i|, (i = 1, \dots, N)$. Utilizing G-L mixed heteroscedastic noise distribution to predict complicated noise-characteristic, from Equation (4), the loss function corresponding to G-L mixed heteroscedastic noise is $l(\xi_i) = \frac{\lambda_1}{2\sigma_i^2} \cdot \xi_i^2 + \lambda_2 \cdot |\xi_i| (i = 1, \dots, N)$. The new model *LSSVR* with G-L mixed heteroscedastic noise-characteristic (*GLMH – LSSVR*) is proposed. The primal problem of *GLMH – LSSVR* is depicted as

$$\begin{aligned} \text{Min}\{g_{P_{GLMH-LSSVR}} &= \frac{1}{2}\omega^T \cdot \omega + \\ &\frac{C}{N} \cdot (\frac{\lambda_1}{2} \cdot \sum_{i=1}^N \frac{1}{\sigma_i^2} \xi_i^2 + \lambda_2 \cdot \sum_{i=1}^N \xi_i)\} \\ \text{s.t. : } \xi_i &= y_i - \omega^T \cdot \Phi(A_i) - b \end{aligned} \tag{10}$$

where the parameter vector is $\omega \in R^n$, $\sigma_i^2 (i = 1, 2, \dots, N)$ are heteroscedastic, and $C > 0$ is the penalty parameter. The weight-factors are $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Proposition 2. *The solution of the primal problem in Equation (10) of GLMH – LSSVR is existent and unique about ω .*

Theorem 2. *The dual problem of model GLMH – LSSVR in Equation (10) is*

$$\begin{aligned} \text{Max}\{g_{D_{GLMH-LSSVR}} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \cdot \alpha_j \cdot K(A_i, A_j) + \\ &\sum_{i=1}^N \alpha_i \cdot y_i - \frac{N}{2C \cdot \lambda_1} \sum_{i=1}^N (\sigma_i^2 \cdot \alpha_i - C \cdot \lambda_2)^2\} \\ \text{s.t. : } \sum_{i=1}^N \alpha_i &= 0 \end{aligned} \tag{11}$$

where $\sigma_i^2 (i = 1, 2, \dots, N)$ are heteroscedastic and $C > 0$ is the penalty parameter. The weight factors are $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Proof. It is easier to derive the proof of Theorem 2 by analogy with Theorem 1. \square

We have

$$\begin{aligned} \omega_i &= \sum_{i=1}^N \alpha_i \cdot \Phi(A_i), \\ b &= \frac{1}{N} \sum_{i=1}^N [y_i - \sum_{j=1}^N \alpha_j \cdot K(A_i, A_j) - \frac{1}{\lambda_1} \cdot (\frac{N \cdot \sigma_i^2 \cdot \alpha_i}{C} - \lambda_2)]. \end{aligned}$$

The decision-maker for GLMH – LSSVR may be expressed as

$$f(A) = \omega^T \cdot \Phi(A) + b = \sum_{i=1}^N \alpha_i K(A_i, A) + b,$$

where the parameter vector is $\omega \in R^n$, $\Phi : R^n \rightarrow H$, and $K(A_i, A_j)$ is the kernel function.

Suppose noise in Equation (2) is G-L mixed-homoscedastic-noise, in which Gaussian-noise of zero mean and homoscedastic-variance σ^2 , Theorem 1 can be deduced from Theorem 2.

4. Solution from ALM

In this section, we use Augmented Lagrange-multiplier method (ALM) [32] to solve the dual problem in Equation (8) by applying Gradient descent or Newton’s method to a sequence of equality-constrained problems. By eliminating equality constraints, arbitrary equality constraints can be reduced to equivalent unconstrained problems [33,34]. If there are large-scale training samples, some rapid optimization techniques can be combined with the proposed model, for example the sequential minimal optimization (SMO) algorithm [29] and the stochastic gradient decent (SDG) algorithm [35].

Theorems 1 and 2 provide effective recognition techniques for GLM – LSSVR and GLMH – LSSVR, respectively. In this section, we derive the solution from ALM and the algorithm for model LSSVR of G-L mixed homoscedastic noise characteristic (GLM – LSSVR). Analogously, the solution of model GLMH – LSSVR can be obtained by ALM method.

- (1) Let dataset be $D_N = \{(A_1, y_1), (A_2, y_2), \dots, (A_N, y_N)\}$, where $A_i \in R^n, y_i \in R, i = 1, \dots, N$.
- (2) The optimal parameters C, λ_1, λ_2 were searched by using the 10-fold cross-validation strategy, and the appropriate kernel function $K(\bullet, \bullet)$ was selected.
- (3) Solve model $GLM - LSSVR$ of the problem in Equation (8), and get the optimal solution $\alpha = (\alpha_1, \dots, \alpha_N)$.
- (4) Build the decision-function as follows

$$f(A) = \omega^T \cdot \Phi(A) + b = \sum_{i=1}^N \alpha_i K(A_i, A) + b.$$

The parameter vector is $\omega \in R^n, b = \frac{1}{N} \sum_{i=1}^N [y_i - \sum_{j=1}^N \alpha_j \cdot K(A_i, A_j) - \frac{1}{\lambda_1} \cdot (\frac{N \cdot \sigma^2 \cdot \alpha_i}{C} - \lambda_2)]$, $\Phi : R^n \rightarrow H, (\Phi(A_i) \cdot \Phi(A_j)) ((i, j = 1, 2, \dots, N))$ is the inner product in $H, K(A_i, A_j) = (\Phi(A_i) \cdot \Phi(A_j))$ is a kernel function.

5. Case Study

This section tests and verifies the validity of constructed model $GLM - LSSVR$ by comparing it with other techniques in the Heilongjiang, China dataset D_N . This case study consists of the following subsections: G-L mixed-noise characteristic of wind speed, prediction performance evaluation criteria, and short-term wind-speed forecasting based on an actual dataset.

5.1. G-L Mixed-Noise-Characteristic of Wind-Speed

To demonstrate the effectiveness of the proposed model, we collected wind speed data from Heilongjiang. The dataset consists of more than one year of wind speed data, recording wind speed values every 10 min. We first discovered the G-L mixed noise and conducted experiments on it. We found that turbulence is the main reason for the high uncertainty of wind speed random fluctuations. From the perspective of wind energy, the most significant feature of wind energy resources is their variability. Now, it shows the distribution of wind speed. Take a wind speed value every 5 s and calculate the histogram of wind speed within 1–2 h. Two typical distributions are given: one is calculated when the wind speed is high and the other is calculated when the wind speed is low (see Figures 2 and 3, respectively).

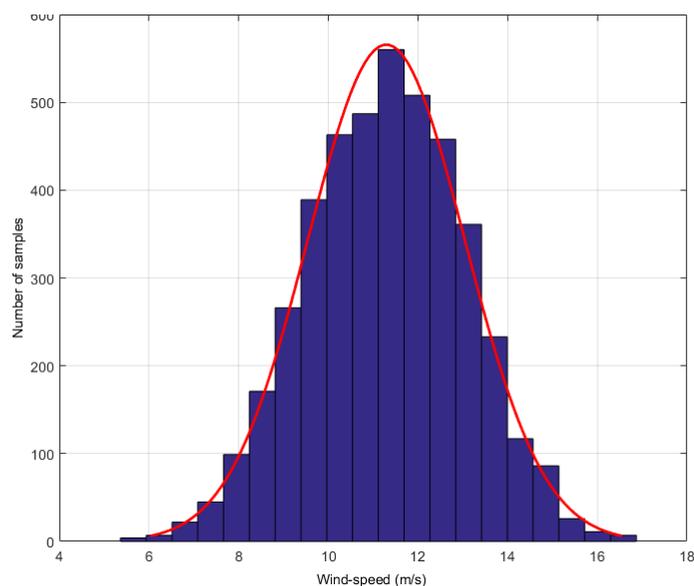


Figure 2. High wind speed distribution.

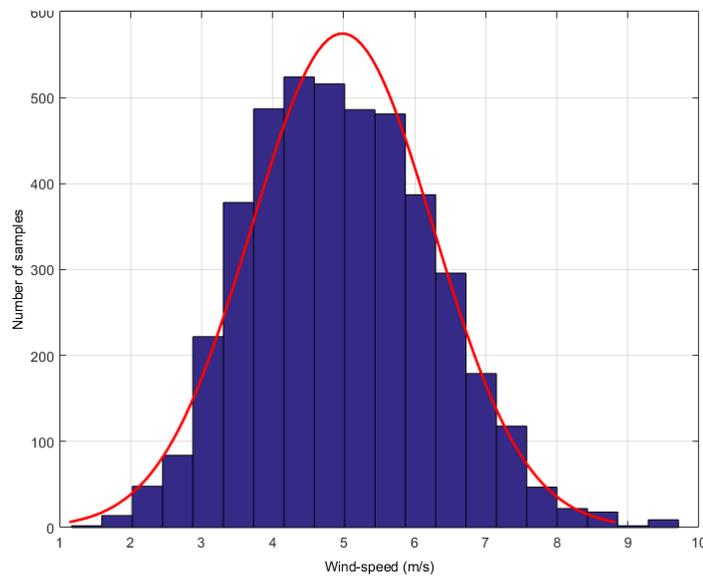


Figure 3. Low wind speed distribution.

We analyzed the one-month time-series dataset, and used the persistence method to investigate the error distribution [32]. The results show that the wind speed error ζ obtained from the persistence prediction is not subject to single distribution, while approximately to G-L mixed distribution, and *PDF* of ζ is $p(\zeta) = \frac{1}{2}e^{-|\zeta|} \cdot \frac{1}{2\sigma^2}\zeta^2$, as shown in Figure 4.

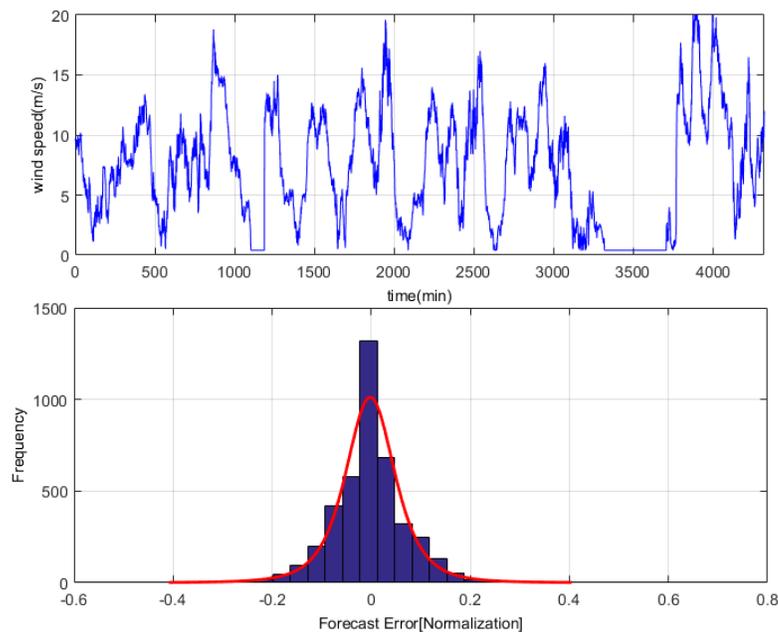


Figure 4. G-L mixed distribution of wind-speed forecasting-error with the persistence method.

As can be seen from the above charts and figures, wind speed error approximately satisfies G-L mixed distribution. This is a mixed kind of task.

5.2. Prediction Performance Evaluation Criteria

It is generally known that no prediction model forecasts perfectly. The predictable performance of ν -SVR, GN-SVR, LSSVR, and GLM-LSSVR also has certain evaluation criteria, for example MAE (mean absolute error), RMSE (root mean square error), MAPE (mean absolute percentage error), and SEP (the standard error of prediction). The four criteria be defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y'_i - y_i|, \quad (12)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y'_i - y_i|}{y_i} \times 100\%, \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y'_i - y_i)^2}, \quad (14)$$

$$SEP = \frac{RMSE}{\bar{y}} \times 100\%, \quad (15)$$

where N is the size of the dataset D_N , y_i is the i th actual observed data, and y'_i is the i th forecasted-result. \bar{y} is the mean value of observations $y_i \in D_N$ [36–40]. MAE shows how similar the predicted value is to the observed value, while RMSE measures overall deviation between predicted value and observed value. MAPE is the ratio between error and observed value. SEP is the ratio of RMSE to average observation. They are dimensionless measurements of accuracy of wind speed system, and are sensitive to small changes.

5.3. Short-Term Wind-Speed Forecasting with Real dataset

In this section, 2160 consecutive data (1–2160, time span of 15-days) are extracted as the training set and 720 consecutive data (2161–2880, time span of 5-days) are extracted as the testing set. The input vector is $\bar{A}_i = (x_{i-11}, x_{i-10}, \dots, x_{i-1}, x_i)$, x_j is the actual observed data of wind speed at moment j ($j = i - 11, i - 10, \dots, i$), and the forecasting value is x_{i+step} , where $step = 1, 3, 6$. That is, the above models are used to forecast wind speed of each point x_i after 10, 30 and 60 min, respectively. Figures 5–13 describe the forecasting results given by models ν -SVR, GN-SVR, LSSVR, and GLM-LSSVR.

The models ν -SVR, GN-SVR, LSSVR, and GLM-LSSVR were implemented in Matlab 7.8. Initial parameters of GLM-LSSVR were $C \in [1, 200]$, $\nu \in (0, 1)$, and $\lambda_1, \lambda_2 \in [0, 1]$. The optimal parameters $C, \nu, \lambda_1, \lambda_2$ were searched by using 10-fold cross-validation technique. The technology of parameter selection is studied in detail in [41,42]. In this simulation, parameters were set to $C = 181, \nu = 0.5, \lambda_1 = 0.5, \lambda_2 = 0.5$. The practical application demonstrates that both polynomial kernel and Gaussian kernel perform well under the assumption of smoothness. Under these circumstances, models ν -SVR, GN-SVR, LSSVR, and GLM-LSSVR employ polynomial and Gaussian kernel functions [43]:

$$K(A_i, A_j) = ((A_i, A_j) + 1)^d,$$

$$K(A_i, A_j) = e^{-\frac{\|A_i - A_j\|^2}{\sigma^2}},$$

where d is a positive integer and σ is a positive number.

The dual problem of ν -SVR and SVR of the Gaussian-noise model (GN-SVR) and LSSVR are as follows.

ν -SVR: The authors of [41,44] define the dual problem of ν -SVR as

$$\begin{aligned}
 & \text{Max}\{g_{D_{\nu-SVR}} = -\frac{1}{2} \sum_{i \in RSV} \sum_{j \in RSV} (\alpha_i^* - \alpha_j)(\alpha_j^* - \alpha_j) \\
 & \quad \cdot K(A_i, A_j) + \sum_{i \in RSV} (\alpha_i^* - \alpha_i) \cdot y_i\} \\
 \text{s.t. : } & \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0 \\
 & 0 \leq \alpha_i^{(*)} \leq \frac{C}{N}, i = 1, \dots, N \\
 & \sum_{i=1}^N (\alpha_i + \alpha_i^*) \leq C \cdot \nu, i = 1, \dots, N.
 \end{aligned}
 \tag{16}$$

GN – SVR: The authors of [45,46] studied SVR with equality constraints and inequality constraints. The loss-function of Gaussian-noise is $c(\zeta_i) = \zeta_i^2/2, (i = 1, \dots, N)$. Thus, the dual problem of GN – SVR is

$$\begin{aligned}
 & \text{max}\{g_{D_{GN-SVR}} = -\frac{1}{2} \sum_{i \in RSV} \sum_{j \in RSV} (\alpha_i^* - \alpha_j)(\alpha_j^* - \alpha_j)K(A_i, A_j) + \\
 & \quad \sum_{i \in RSV} (\alpha_i^* - \alpha_i)y_i - \frac{N}{2C} \sum_{i=1}^N (\alpha_i^2 + (\alpha_i^*)^2)\} \\
 \text{s.t. : } & \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0 \\
 & 0 \leq \alpha_i^{(*)} \leq \frac{C}{N}, i = 1, \dots, N \\
 & \sum_{i=1}^N (\alpha_i + \alpha_i^*) \leq C \cdot \nu, i = 1, \dots, N.
 \end{aligned}
 \tag{17}$$

LSSVR: [22] studied LS – SVR for Gaussian-noise model. The dual problem of LS – SVR is

$$\begin{aligned}
 & \text{max}\{g_{D_{LS-SVR}} = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i \cdot \alpha_j \cdot K(A_i, A_j)) \\
 & \quad + \sum_{i=1}^N (\alpha_i \cdot y_i) - \frac{N}{2C} \cdot \sum_{i=1}^N (\alpha_i^2)\} \\
 \text{s.t. : } & \sum_{i=1}^N \alpha_i = 0.
 \end{aligned}
 \tag{18}$$

where $\zeta_i, \tilde{\zeta}_i^*$ are slack-variables. $C > 0, \nu \in (0, 1]$ are constants. For $\nu - SVR$ and $GN - SVR$, the size of c is not gained, but is a variable whose value is compromised by a constant with the model complexity and relaxation variables through ν [35].

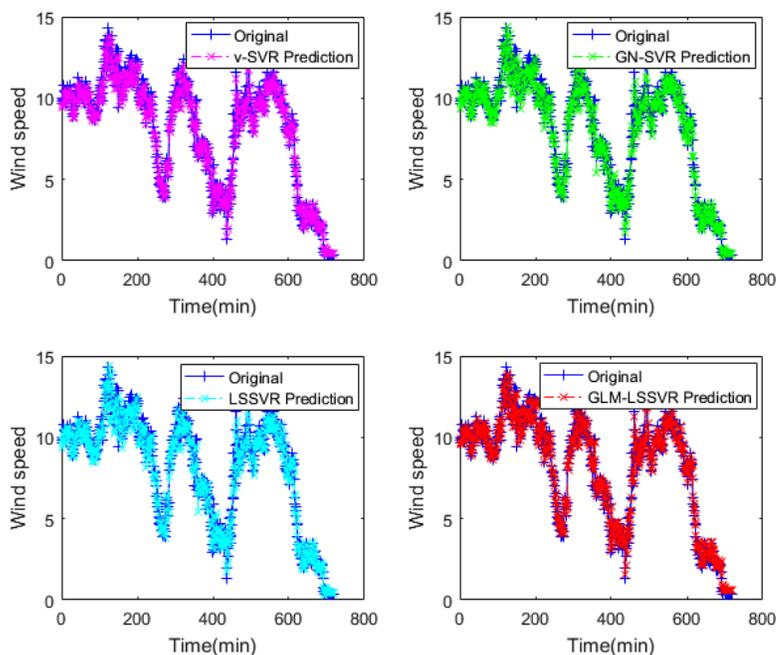


Figure 5. Result of four wind-speed forecasting models after 10 min.

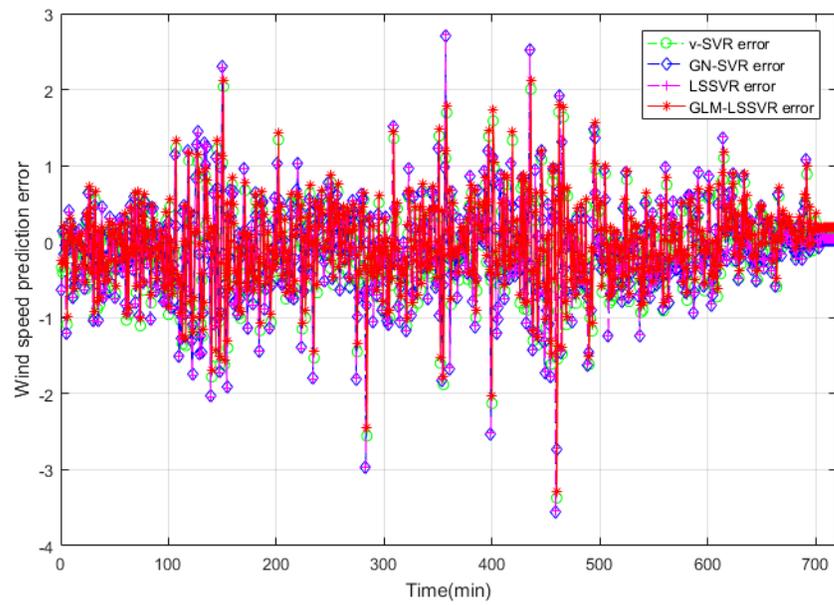


Figure 6. Error of four wind-speed forecasting models after 10 min.

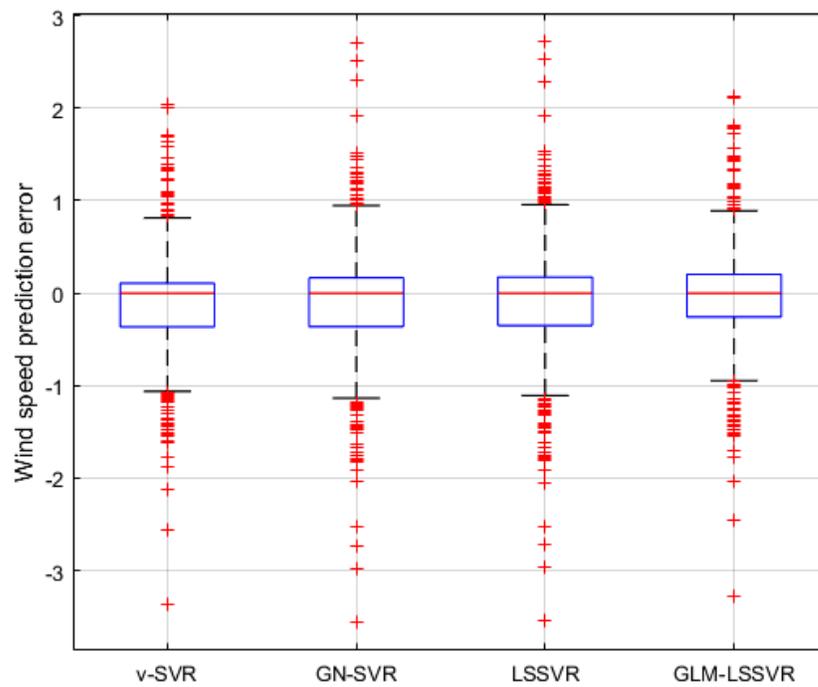


Figure 7. Residual box plot of four wind-speed forecasting models after 10 min.

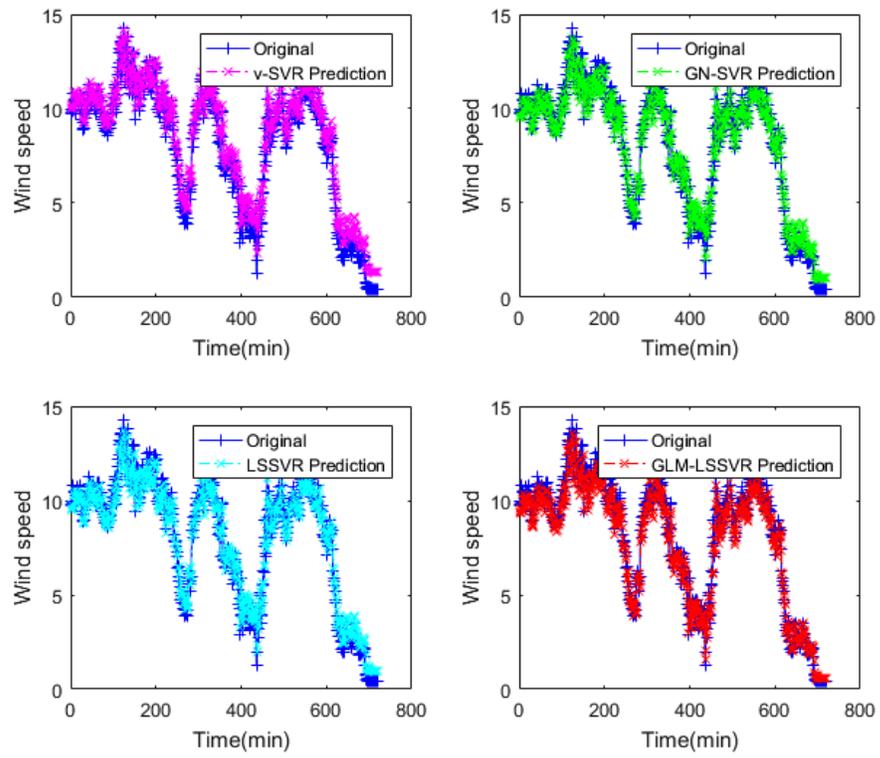


Figure 8. Result of four wind-speed forecasting models after 30 min.

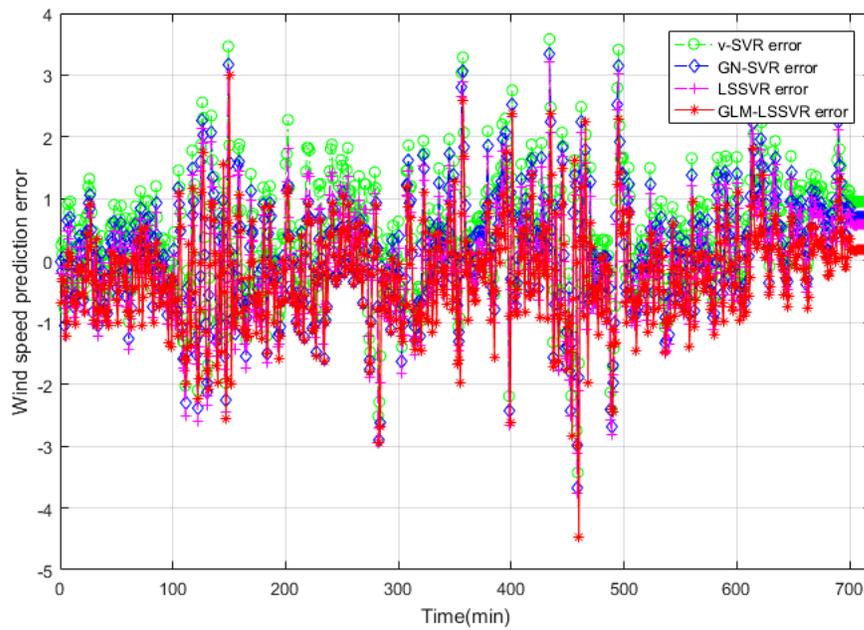


Figure 9. Error of four wind-speed forecasting models after 30 min.

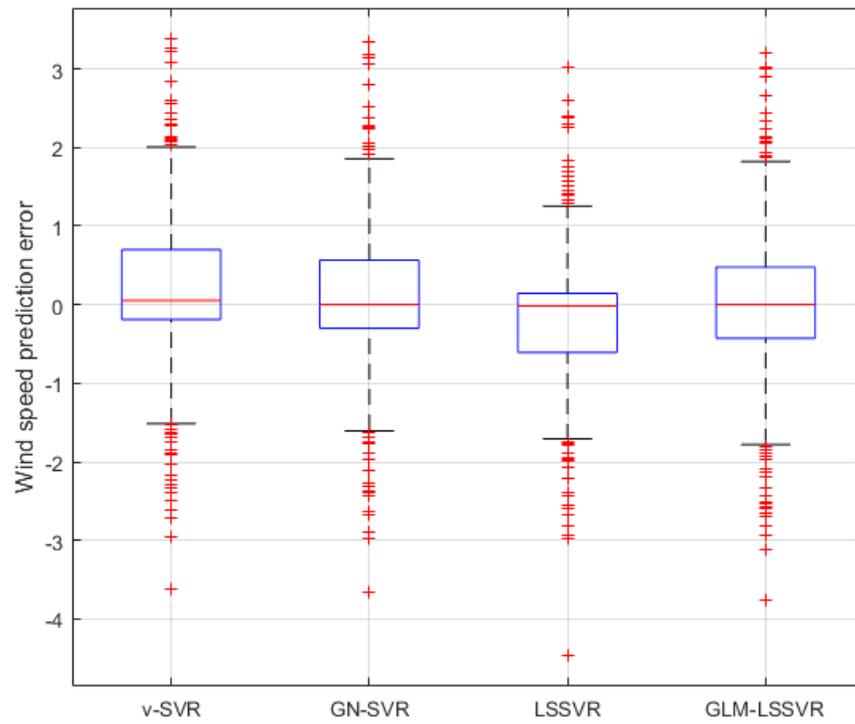


Figure 10. Residual box plot of four wind-speed forecasting models after 30 min.

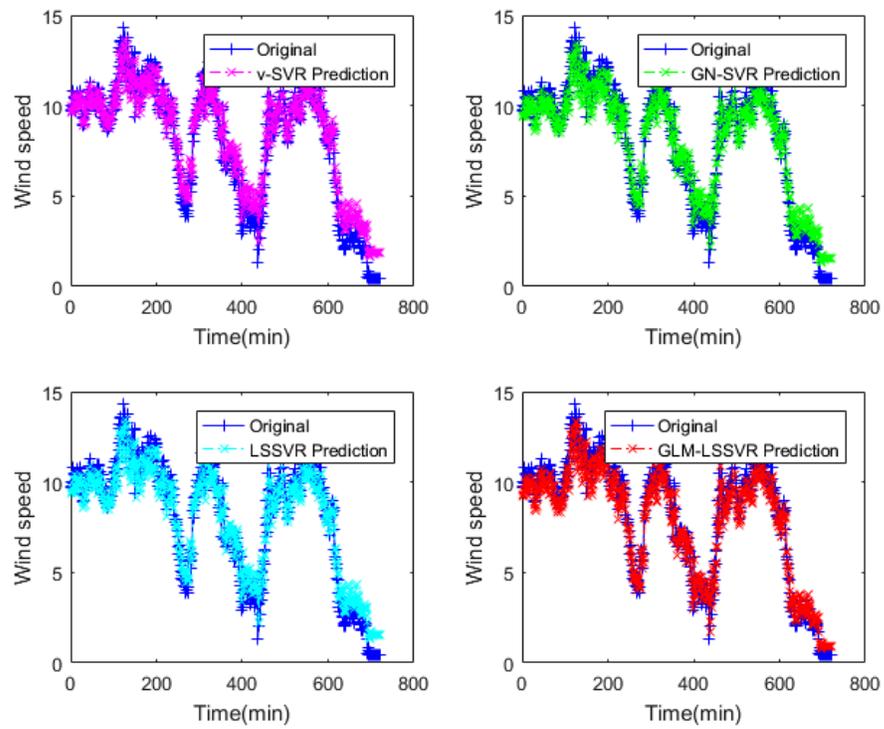


Figure 11. Result of four wind-speed forecasting models after 60 min.

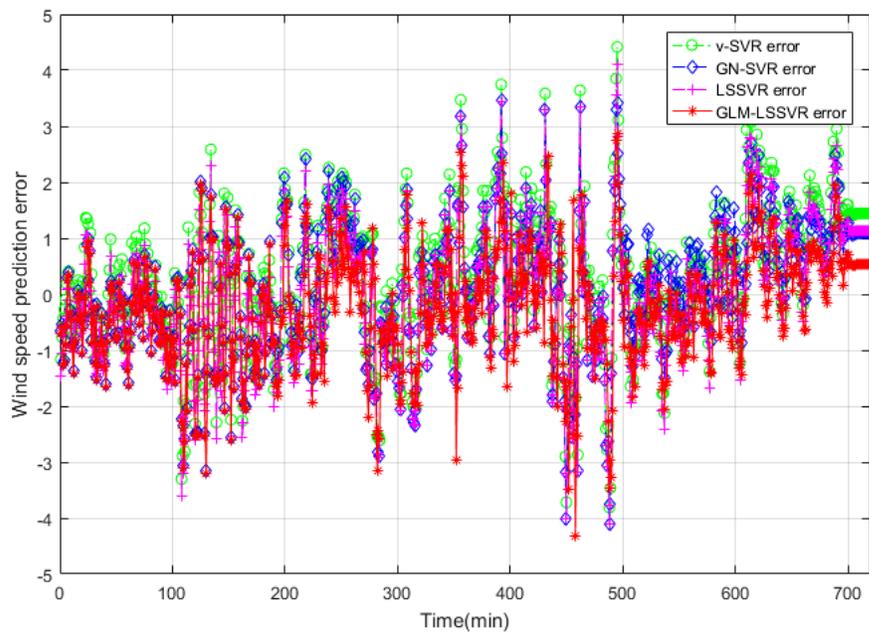


Figure 12. Error of four wind-speed forecasting models after 60 min.

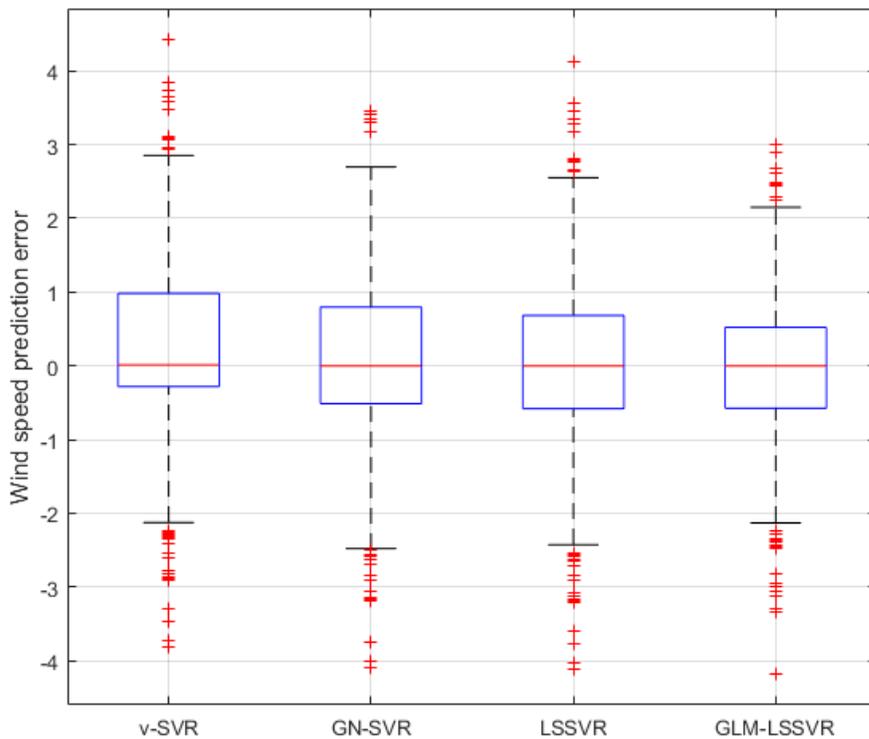


Figure 13. Residual box plot of four wind-speed forecasting models after 60 min.

In Figures 5, 8 and 11, wind-speed forecasting-results at A_i -point of ν – SVR, GN – SVR, LSSVR, and GLM – LSSVR are presented after 10, 30, and 60 min, respectively. Figures 6, 9, and 12 show the

error statistic of wind-speed prediction using the above four models. The box plots (Figures 7, 10, and 13) of several noise levels further intuitively demonstrate the comparative effect of error statistics using the above four wind-speed forecasting models. The statistical criteria of *MAE*, *MAPE*, *RMSE* and *SEP* are displayed in Tables 1–3.

Table 1. Error statistic of four wind-speed forecasting models after 10 min.

Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	SEP (%)
$\nu - SVR$	0.4280	0.5833	8.02	7.12
$GN - SVR$	0.4256	0.5789	7.92	7.07
$LSSVR$	0.4219	0.5768	7.94	7.06
$GLM - LSSVR$	0.4190	0.5711	7.91	7.05

Table 2. Error statistic of four wind-speed forecasting models after 30 min.

Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	SEP (%)
$\nu - SVR$	0.7979	1.0116	23.36	12.53
$GN - SVR$	0.7368	0.9886	19.93	11.89
$LSSVR$	0.7109	0.9226	17.17	11.43
$GLM - LSSVR$	0.6185	0.8241	10.71	10.19

Table 3. Error statistic of four wind-speed forecasting models after 60 min.

Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	SEP (%)
$\nu - SVR$	0.9994	1.2580	33.93	15.66
$GN - SVR$	0.9728	1.2355	31.78	15.37
$LSSVR$	0.9646	1.2177	29.01	15.16
$GLM - LSSVR$	0.8835	1.1180	25.72	13.97

From box-whisker plots in Figures 7, 10, and 13, as well as Tables 1–3, it can be concluded that, in most cases, the forecasting-error of $GLM - LSSVR$ calculation is superior to $\nu - SVR$, $GN - SVR$ and $LSSVR$. With the increase of prediction horizon to 30 and 60 min, the forecasting error of different models increases and the relative error decreases. Thus, in these cases, it is not that important. However, Tables 1–3 show that, under all the criteria of *MAE*, *MAPE*, *RMSE*, and *SEP*, the Gaussian–Laplacian mixed-noise model is slightly better than the classical model.

6. Conclusions

Most existing regression-techniques suppose that the noise model is single. Wind-speed forecasting is complicated due to its volatility and uncertainty, thus it is difficult to model with a single-noise distribution. This section summarizes our main work: (1) optimal empirical risk loss of G-L mixed noise is deduced by Bayesian principle; (2) the $LSSVR$ of G-L mixed homoscedastic noise ($GLM - LSSVR$) and G-L mixed heteroscedastic noise ($GLMH - LSSVR$) for complicate noise is developed; (3) the dual problem of $GLM - LSSVR$ and $GLMH - LSSVR$ is obtained using Lagrange-functional and according to KKT conditions; (4) the stability and effectiveness of the algorithm are guaranteed by solving $GLM - LSSVR$

with the ALM method; and (5) the proposed technology is used to predict short-term wind speed by historical data, and then forecast the wind speed at some time after 10, 30, and 60 min, respectively. The comparison results display that the proposed model is better than classical technologies in statistical criteria.

In the same way, we can also study Gaussian–Laplacian, or Gaussian–Weibull mixed noise classification models. The new hybrid noise models would effectively solve complicated noise classification problems.

Author Contributions: Conceptualization, S.Z.; Formal analysis, S.Z. and T.Z.; Methodology, S.Z. and L.S.; Writing–original draft, S.Z. and T.Z.; Writing–review & editing, W.W. and B.C. All authors have read and approved the final published version. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National natural science foundation of China (NSFC) (Nos. 11702087 and 61772176) and Natural Science Foundation Project of Henan (No. 182300410130).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LR	Linear regression model
ν -SVR	ν -Support vector regression
GN-SVR	ν -SVR model of Gaussian homoscedastic-noise
LSSVR	Least squares support vector regression model
GLM-LSSVR	LSSVR model of Gaussian–Laplacian mixed homoscedastic-noise
ALM	Augmented Lagrange multiplier method

References

1. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
2. Tikhonov, A.A.; Arsenin, V.Y. *Solutions of Ill-Posed Problems*; New York Wiley: New York, NY, USA, 1977.
3. Gonen A.; Orabona, F.; Shalev-Shwartz, S. Solving Ridge Regression using Sketched Preconditioned SVRG. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016.
4. Hoerl, A.E. Application of ridge analysis to regression problems. *Chem. Eng. Prog.* **1962**, *58*, 54–59.
5. Zhang, Z.H.; Dai, G.; Xu, C.F. Regularized Discriminant Analysis, Ridge Regression and Beyond. *J. Mach. Learn. Res.* **2010**, *11*, 2199–2228.
6. Sun, L.; Wang, L.; Ding, W.; Qian, Y.; Xu, J. Feature Selection Using Fuzzy Neighborhood Entropy-Based Uncertainty Measures for Fuzzy Neighborhood Multigranulation Rough Sets. *IEEE Trans. Fuzzy Syst.* **2020**. [[CrossRef](#)]
7. Jiao, L.C.; Bo, L.F.; Wang, L. Fast Sparse Approximation for Least Squares Support Vector Machine. *IEEE Trans. Neural Netw.* **2007**, *18*, 685–697. [[CrossRef](#)]
8. Völgyesi, L.; Palánc, B.; Fekete, K.; Popper, G. Application of Kernel Ridge Regression to Network Levelling via Mathematica. *Geophys. Res. Abstr.* **2005**, *73*, 263–276.
9. Sun, L.; Zhang, X.; Qian, Y.; Xu, J.; Zhang, S. Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification. *Inf. Sci.* **2019**, *502*, 18–41. [[CrossRef](#)]
10. Douak, F.; Melgani, F.; Benoudjit, N. Kernel ridge regression with active learning for wind-speed prediction. *Appl. Energy.* **2013**, *103*, 328–340. [[CrossRef](#)]
11. Alexiadis, M.C.; Dokopoulos, P.S.; Sahsamanoglou, H.S.; Manousaridis, I.M. Short term forecasting of wind speed and related electrical power. *J. Sol. Energy* **1998**, *63*, 61–68. [[CrossRef](#)]
12. Negnevitsky, M.; Potter, C.W. Innovative short-term wind generation prediction techniques. In Proceedings of the Power Systems Conference and Exposition, Atlanta, GA, USA, 29 October–1 November 2006.
13. Torres, J.L.; Garcia, A.; De Blas, M.; De Francisco, A. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *J. Sol. Energy* **2005**, *79*, 65–77. [[CrossRef](#)]
14. Kavasseri, R.G.; Seetharaman, K. Day-ahead wind-speed forecasting using f-ARIMA models. *Renew. Energy* **2009**, *34*, 1388–1393. [[CrossRef](#)]

15. Li, G.; Shi, J. On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy* **2010**, *87*, 2313–2320. [\[CrossRef\]](#)
16. Hu, Q.; Zhang, R.; Zhou, Y. Transfer learning for short-term wind-speed prediction with deep neural networks. *Renew. Energy* **2016**, *85*, 83–95. [\[CrossRef\]](#)
17. Salcedo-Sanz, S.; Ortiz-García, E.G.; Pérez-Bellido, Á.M.; Portilla-Figueras, A.; Prieto, L. Short term wind-speed prediction based on evolutionary support vector regression algorithms. *Expert Syst. Appl.* **2011**, *38*, 4052–4057. [\[CrossRef\]](#)
18. Zhou, J.; Shi, J.; Li, G. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers. Manag.* **2011**, *52*, 1990–1998. [\[CrossRef\]](#)
19. Liu, H.; Tian, H.-Q.; Chen, C.; Li, Y.-F. A hybrid statistical method to predict wind speed and wind power. *Renew. Energy* **2010**, *35*, 1857–1861. [\[CrossRef\]](#)
20. Wang, Y.; Hu, Q.; Li, L.; Foley, A.M.; Srinivasan, D. Approaches to wind power curve modeling: A review and discussion. *Renew. Sustain. Energy Rev.* **2019**, *116*, 109422. [\[CrossRef\]](#)
21. Suykens, J.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [\[CrossRef\]](#)
22. Suykens, J.; Lukas, L.; Vandewalle, J. Sparse approximation using least square vector machines. In Proceedings of the IEEE International Symposium on Circuits and Systems, Geneva, Switzerland, 28–31 May 2000; pp. 757–760.
23. Suykens, J.; De Brabanter, J.; Lukas, L.; Vandewalle, J. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing* **2002**, *48*, 85–105. [\[CrossRef\]](#)
24. Du, P.; Wang, J.; Guo, Z.; Yang, W. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. *Energy Convers. Manag.* **2017**, *150*, 90–107. [\[CrossRef\]](#)
25. Sun, L.; Wang, L.; Ding, W.; Qian, Y.; Xu, J. Neighborhood multi-granulation rough sets-based attribute reduction using Lebesgue and entropy measures in incomplete neighborhood decision systems. *Knowl.-Based Syst.* **2020**, *192*, 105373. [\[CrossRef\]](#)
26. Jiang, Y.; Huang, G.Q. A hybrid method based on singular spectrum analysis, firefly algorithm, and BP neural network for short-term wind-speed forecasting. *Energies* **2016**, *9*, 757.
27. Jiang, Y.; Huang, G. Short-term wind speed prediction: Hybrid of ensemble empirical mode decomposition, feature selection and error correction. *Energy Convers. Manag.* **2017**, *144*, 340–350. [\[CrossRef\]](#)
28. Zhang, S.; Zhou, T.; Sun, L.; Wang, W.; Wang, C.; Mao, W. ν -Support Vector Regression Model Based on Gauss-Laplace Mixture Noise Characteristic for Wind Speed Prediction. *Entropy* **2019**, *21*, 1056. [\[CrossRef\]](#)
29. Shevade, S.; Keerthi, S.S.; Bhattacharyya, C.; Murthy, K. Improvements to the SMO algorithm for SVM regression. *IEEE Trans. Neural Netw.* **2000**, *11*, 1188–1193. [\[CrossRef\]](#)
30. Klaus-Robert, M.; Sebastian, M. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw.* **2001**, *12*, 181–202.
31. Chu, W.; Keerthi, S.; Ong, C.J. Bayesian Support Vector Regression Using a Unified Loss Function. *IEEE Trans. Neural Netw.* **2004**, *15*, 29–44. [\[CrossRef\]](#)
32. Rockafellar, R.T. Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming. *SIAM J. Control* **1974**, *12*, 268–285. [\[CrossRef\]](#)
33. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004; pp. 521–620.
34. Wang, S.; Zhang, N.; Wu, L.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [\[CrossRef\]](#)
35. Bordes, A.; Bottou, L.; Gallinari, P. SGD-QN: Careful quasiNewton stochastic gradient descent. *J. Mach. Learn. Res.* **2009**, *10*, 1737–1754.
36. Bludszweit, H.; Dominguez-Navarro, J.; Llombart, A. Statistical Analysis of Wind Power Forecast Error. *IEEE Trans. Power Syst.* **2008**, *23*, 983–991. [\[CrossRef\]](#)
37. Fabbri, A.; Román, T.G.S.; Abbad, J.R.; Quezada, V.H.M. Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. *IEEE Trans. Power Syst.* **2005**, *20*, 1440–1446. [\[CrossRef\]](#)
38. Guo, Z.; Zhao, J.; Zhang, W.; Wang, J. A corrected hybrid approach for wind speed prediction in Hexi Corridor of China. *Energy* **2011**, *36*, 1668–1679. [\[CrossRef\]](#)

39. Wang, J.Z.; Hu, J.M. A robust combination approach for short-term wind-speed forecasting and analysis-Combination of the ARIMA, ELM, SVM and LSSVM forecasts using a GPR model. *Energy* **2015**, *93*, 41–56. [[CrossRef](#)]
40. Abdoos, A.A. A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing* **2016**, *203*, 111–120. [[CrossRef](#)]
41. Chalimourda, A.; Schölkopf, B.; Smola, A.J. Experimentally optimal ν in support vector regression for different noise models and parameter settings. *Neural Netw.* **2004**, *17*, 127–141. [[CrossRef](#)]
42. Cherkassky, V.; Ma, Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw.* **2004**, *17*, 113–126. [[CrossRef](#)]
43. Kwok, J.T.; Tsang, I.W. Linear dependency between and the input noise in ϵ -support vector regression. *IEEE Trans. Neural Netw.* **2003**, *14*, 544–553. [[CrossRef](#)]
44. Schölkopf, B.; Smola, A.J.; Williamson, R.C.; Bartlett, P. New Support Vector Algorithms. *Neural Comput.* **2000**, *12*, 1207–1245.
45. Wu, Q. A hybrid-forecasting model based on Gaussian support vector machine and chaotic particle swarm optimization. *Expert Syst. Appl.* **2010**, *37*, 2388–2394. [[CrossRef](#)]
46. Wu, Q.; Law, R. The forecasting model based on modified SVRM and PSO penalizing Gaussian noise. *Expert Syst. Appl.* **2011**, *38*, 1887–1894. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).